Tomasi-Kanade Factorization Method

CMPE 264: Image Analysis and Computer Vision
Hai Tao
3D reconstruction with the orthographic camera model

Problem statement
- Assumption: an orthographic camera
- Given: point correspondences across multiple views
- Recover: camera motion and 3D coordinates for feature points

We will describe a method described in Tomasi and Kanade’92. This method has been extended to deal with other affine cameras such as weak perspective cameras and paraperspective cameras. Based on the same principal, algorithms have also been proposed for 3D reconstruction for perspective cameras and scenes with non-rigid motions.

Reference:
- Forsyth and Ponce book: Section 14.3
Orthographic camera model

Recall that for an orthographic camera, the projection matrix is

\[
M = \begin{bmatrix}
    s_x & 0 & u_0 \\
    0 & s_y & v_0 \\
    0 & 0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
    r_1^T \\
    r_2^T \\
    0 \\
\end{bmatrix}
\begin{bmatrix}
    -r_1^T r_2^T \\
    1 \\
\end{bmatrix}
\]

Since the last row is [0,0,1], the projection process can be simplified as

\[
p = \begin{bmatrix}
    s_x & 0 \\
    0 & s_y \\
\end{bmatrix}
\begin{bmatrix}
    r_1^T \\
    r_2^T \\
\end{bmatrix}
\begin{bmatrix}
    P - r_1^T r_2^T + [u_0] \\
\end{bmatrix}
\]

where \( p = [x, y]^T \) and \( P = [X, Y, Z]^T \). If we assume the camera intrinsic parameters \((s_x, s_y, u_0, v_0)\) are known, then the image coordinates can be converted into the image plane coordinates system as

\[
p' = (p - [u_0, v_0]^T) \begin{bmatrix}
    1/s_x & 0 \\
    0 & 1/s_y \\
\end{bmatrix}
\]

For the \( i \)th feature points in the \( j \)th image, we will use \( p_i^{(j)} = [x_i^{(j)}, y_i^{(j)}]^T \) to denote its image plane coordinates.
Registered measurement matrix – remove the centroid

- \( P_i \) and its projection in the \( j \)th image \( p_i^{(j)} \) are then related by

\[
p_i^{(j)} = \begin{bmatrix}
r_1^{(j)T} \\
r_2^{(j)T}
\end{bmatrix} P_i + \begin{bmatrix}
-r_1^{(j)T}T^{(j)} \\
-r_2^{(j)T}T^{(j)}
\end{bmatrix}
\]

We can put images plane coordinates of \( M \) feature points together as

\[
\begin{bmatrix}
p_1^{(j)} & \cdots & p_M^{(j)}
\end{bmatrix} = \begin{bmatrix}
r_1^{(j)T} \\
r_2^{(j)T}
\end{bmatrix} \begin{bmatrix}
P_1 & \cdots & P_M
\end{bmatrix} + \begin{bmatrix}
-r_1^{(j)T}T^{(j)} \\
-r_2^{(j)T}T^{(j)}
\end{bmatrix}
\]

(1)

Since for orthographic projection, the centroid \( \bar{P} \) of the \( N \) scene points is mapped to the centroid of the \( N \) image feature points \( \bar{p} \). Therefore,

\[
\bar{p}^{(j)} = \begin{bmatrix}
r_1^{(j)T} \\
r_2^{(j)T}
\end{bmatrix} \bar{P} + \begin{bmatrix}
-r_1^{(j)T}T^{(j)} \\
-r_2^{(j)T}T^{(j)}
\end{bmatrix}
\]

(2)
Rank theorem

As the result,

\[
\begin{bmatrix}
    p_1^{(j)} - \bar{p}^{(j)} & \ldots & p_M^{(j)} - \bar{p}^{(j)}
\end{bmatrix} =
\begin{bmatrix}
    r_1^{(j)T} \\
    r_2^{(j)T}
\end{bmatrix}
\begin{bmatrix}
    P_1 - \bar{P} & \ldots & P_M - \bar{P}
\end{bmatrix}
\]

W can combine image correspondences from N images to form the registered measurement matrix W

\[
W =
\begin{bmatrix}
    p_1^{(1)} - \bar{p}^{(1)} & \ldots & p_M^{(1)} - \bar{p}^{(1)} \\
    \vdots & \ddots & \vdots \\
    p_1^{(N)} - \bar{p}^{(N)} & \ldots & p_M^{(N)} - \bar{p}^{(N)}
\end{bmatrix} =
\begin{bmatrix}
    r_1^{(1)T} \\
    r_2^{(1)T} \\
    \vdots \\
    r_1^{(N)T} \\
    r_2^{(N)T}
\end{bmatrix}
\begin{bmatrix}
    P_1 - \bar{P} & \ldots & P_M - \bar{P}
\end{bmatrix} = RS
\]

Rank Theorem: Without the noise, the registered measure matrix W is at most rank three (Because the motion matrix R and the structure matrix S are rank three)
Reconstruction up to a 3 by 3 transformation

With noise, the rank of the registered measurement matrix may not be three. This problem, however, can be solved by keeping only the first three greatest eigenvalues in the singular value decomposition of \( W \). More specifically, if the SVD of \( W \) is \( W = UDV^T \) and

\[
D = \begin{bmatrix}
\lambda_1 & & \\
& \ddots & \\
& & \lambda_M \\
0 & & \\
\end{bmatrix}
\]

then \( W' = U'D'V'^T \), or

\[
W' = [U_1 \ U_2 \ U_3] \begin{bmatrix}
\lambda_1 & 0 & V_1^T \\
\lambda_2 & V_2^T \\
0 & \lambda_3 & V_3^T \\
\end{bmatrix}
\]

We assign \( \hat{R} = U'D' \) and \( \hat{S} = V'^T \). This solution is up to a 3 by 3 rank three matrix \( Q \), because \( \hat{R}' = \hat{R}Q \) and \( \hat{S}' = Q^{-1}\hat{S} \) is also a solution.
The metric constraints

From the structure of the $R$ matrix, we notice that addition conditions need to be satisfied. These conditions are

\[
\hat{r}_{2i-1}^T \hat{r}_{2i-1}' = \hat{r}_{2i-1}^T QQ^T \hat{r}_{2i-1} = 1
\]

\[
\hat{r}_{2i}^T \hat{r}_{2i}' = \hat{r}_{2i}^T QQ^T \hat{r}_{2i} = 1
\]

\[
\hat{r}_{2i}^T \hat{r}_{2i-1}' = \hat{r}_{2i}^T QQ^T \hat{r}_{2i-1} = 0
\]

where $i=1,\ldots,N$, $\hat{r}_k$ is the $k$th row of $\hat{R}$

To solve for $Q$, there are two methods

- Solve for $B$ first, where $B = QQ^T$, using a linear method. The $Q$ can be recovered via the Cholesky decomposition. One problem of this method is that the linear solution of $B$ may not be positive definite.
- Directly solve these equations via non-linear least squares methods.
The rotation ambiguity

Once $Q$ is estimated, we obtained the solution $\hat{R}' = \hat{R}Q$ and $\hat{S}' = Q^{-1}\hat{S}$ where $\hat{R}'$ satisfies the metric constraints.

The solution is still up to a rotation matrix because, for any rotation matrix $R_0$, $\hat{R}'' = \hat{R}'R_0$ and $\hat{S}'' = R_0^T\hat{S}'$ is also a solution. To eliminate this ambiguity, we can determine $R_0$ by forcing the first motion matrix as

$$
\begin{bmatrix}
\hat{r}'_{(1)T} \\
\hat{r}'_{(2)T}
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix}
$$

directly

therefore $R_0 = \begin{bmatrix}
\hat{r}'_{(1)} \\
\hat{r}'_{(2)} \\
\hat{r}'_{(1)} \times \hat{r}'_{(2)}
\end{bmatrix}$

Finally, we obtain the solution $W \approx \hat{R}''\hat{S}''$
The structure and the motion

- The recovered structure is represented in the coordinate system with the centroid as the origin.
- The rotation matrix of the \( j \)th camera is
  \[
  R^{(j)} = \begin{bmatrix}
  r^{\ast(j)}_1 & r^{\ast(j)}_2 & r^{\ast(j)}_3 \\
  \end{bmatrix}
  \]
- To obtain the translation of the \( j \)th camera, observe that
  \[
  \overline{P}^{(j)} = \begin{bmatrix}
  r^{\ast(j)T}_1 \\
  r^{\ast(j)T}_2 \\
  \end{bmatrix} [0] + \begin{bmatrix}
  -r^{\ast(j)T}_1 T^{(j)} \\
  -r^{\ast(j)T}_2 T^{(j)} \\
  \end{bmatrix} = \begin{bmatrix}
  -r^{\ast(j)T}_1 T^{(j)} \\
  -r^{\ast(j)T}_2 T^{(j)} \\
  \end{bmatrix}
  \]
  which yields
  \[
  T^{(j)} = \begin{bmatrix}
  r^{\ast(j)}_1 & r^{\ast(j)}_2 & r^{\ast(j)}_3 \\
  \end{bmatrix}
  \begin{bmatrix}
  \overline{P}^{(j)} \\
  \alpha \\
  \end{bmatrix}
  \]
  where \( \alpha \) is an arbitrary real number.
Summary of the algorithm

- Tomasi-Kanade factorization algorithm
  - Obtain the registered measurement matrix \( W \)
  - Compute the SVD of \( W \) as \( W = UDV^T \) and keep the first three largest eigenvalues and obtain \( W' = U' D' V'^T \). Define \( \hat{R} = U' D' \) and \( \hat{S} = V' \)
  - Compute \( Q \) so that \( \hat{R}' = \hat{R} Q \) satisfies the following metric constrains and define \( \hat{S}' = Q^{-1} \hat{S} \)

\[
\hat{r}_{2i-1}^T Q Q^T \hat{r}_{2i-1} = 1 \quad \hat{r}_{2i}^T Q Q^T \hat{r}_{2i} = 1 \quad \hat{r}_{2i}^T Q Q^T \hat{r}_{2i-1} = 0
\]

- Compute \( R_0 = \begin{bmatrix} \hat{r}_1^{(1)} & \hat{r}_2^{(1)} & \hat{r}_1^{(1)} \times \hat{r}_2^{(1)} \end{bmatrix} \) and define \( \hat{R}'' = \hat{R}' R_0 \) and \( \hat{S}'' = R_0^T \hat{S}' \)
- Output structure as \( \hat{S}'' \) and motion for the \( j \)th camera as

\[
R^{(j)} = \begin{bmatrix} r_1^{*(j)} & r_2^{*(j)} & r_3^{*(j)} \end{bmatrix} \\
T^{(j)} = \begin{bmatrix} r_1^{*(j)} & r_2^{*(j)} & r_3^{*(j)} \end{bmatrix} \begin{bmatrix} \bar{P}^{(j)} \\ \alpha \end{bmatrix}
\]

where \( \alpha \) is an arbitrary real number
Experimental results

Reprinted from Tomasi and Kanade 1992